Resonant tunneling and intrinsic bistability in twisted graphene structures

J. F. Rodriguez-Nieva[1], M. S. Dresselhaus[1,2], L. S. Levitov[1]

[1]*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA and*
[2]*Department of Electrical Engineering and Computer Science,*
Massachusetts Institute of Technology, Cambridge, MA 02139, USA

We predict that vertical transport in heterostructures formed by twisted graphene layers can exhibit a unique bistability mechanism. Intrinsically bistable *I-V* characteristics arise from resonant tunneling and interlayer charge coupling, enabling multiple stable states in the sequential tunneling regime. We consider a simple trilayer architecture, with the outer layers acting as the source and drain and the middle layer floating. Under bias, the middle layer can be either resonant or non-resonant with the source and drain layers. The bistability is controlled by geometric device parameters easily tunable in experiments. The nanoscale architecture can enable uniquely fast switching times.

I. INTRODUCTION

Nanoscale systems that can switch between distinct macroscopic states upon variation of some control parameter are in high demand in diverse areas of nanoscience research. Bistable electronic systems which exhibit fast switching are of interest for applications such as low-power memory and logic [1]. Recently, new realizations of intrinsically bistable system have been discovered, both in graphene [2–6] and in other systems [7–9]. In particular, van der Waals heterostructures comprising graphene layers sandwiched between insulating hexagonal boron-nitride (hBN) layers afford electronic environments with tailored band structures and transport characteristics [10]. It was demonstrated that introducing a twist between adjacent graphene layers in such heterostructures can result in a resonant behavior of the tunneling current and non-monotonic *I-V* characteristics [11]. It is therefore tempting to exploit twisted graphene multi-layer structures as a platform for bistable and hysteretic nanoscale systems.

Here we predict intrinsic bistability and hysteretic *I-V* characteristics for vertical transport in heterostructures formed by graphene monolayers separated by hBN barriers, in a twisted arrangement similar to that described in Ref.[11]. Essential for our bistability mechanism are resonances originating from momentum-conserving tunneling between linearly dispersing Dirac bands [12] and occurring when the bands are aligned [11] (see Fig.1b,c). Bistability arises due to current-induced charge accumulation producing an interlayer bias that tunes the interband tunneling in and out of resonance.

Below we focus on the simplest case of a two-step sequential tunneling in a device comprising three graphene monolayers. Such trilayer architecture, pictured in Fig.1a, with the top and bottom layers acting as a source and drain and the middle layer electrically decoupled (floating), is similar to previously studied double-barrier quantum-well (QW) structures [13]. However, our bistability mechanism, originating from resonant tunneling between Dirac bands in graphene layers, is distinct from

that in the QW structures [13]. In our case, multiple stable states arise because the decoupled layer can, for a fixed external bias, be either in a resonant (low resistance) or a non-resonant (high resistance) state. This behavior is illustrated in Fig.1c.

The bistability is governed by geometric parameters – the twist angle θ and the interlayer distances d_{ij} – which are easily tunable in experiments. The twist angle
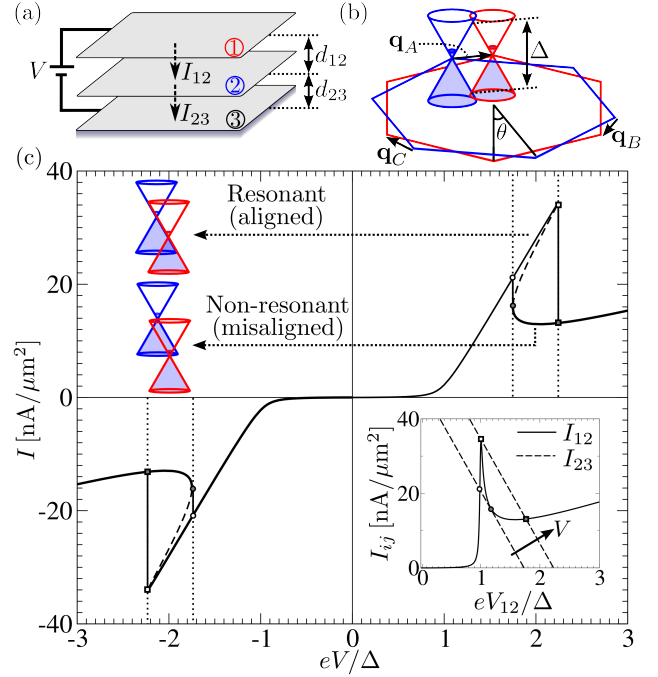

FIG. 1. (a) Trilayer graphene heterostructure schematics, with layers labeled 1 to 3. Here I_{ij} and d_{ij} are the interlayer currents and distances. (b) Band structure of the twisted graphene layers 1(blue) and 2(red). The twist angle θ defines a characteristic energy Δ [Eq.(1)] and three superlattice wave vectors $\mathbf{q}_{A,B,C}$ [Eq.(12)]. (c) Bistable *I-V* characteristics. The resonant and non-resonant bistable states are illustrated in the top left inset (details are discussed in Fig.3). The procedure for finding bistable solutions is illustrated in the bottom right inset [see Eq.(9) and accompanying discussion].

controls the Dirac cones' displacement in the two layers and the energy at which the cones intersect (see Fig.1b),

$$|\mathbf{q}_A| = (8\pi/3a_0)\sin(\theta/2), \quad \Delta = \hbar v_F |\mathbf{q}_A|, \qquad (1)$$

where $v_F \approx 10^6$ m/s is the carrier velocity and $a_0 \approx 2.46$ Å is the graphene lattice constant. The distances d_{ij}, marked in Fig.1a, determine the interlayer tunnel conductance values $G_{ij} \sim e^{2d_{ij}/\lambda}$, where λ is the WKB length governing the tunneling amplitude dependence on barrier width. In what follows we will use the conductance ratio

$$Z = G_{12}/G_{23} \sim e^{2(d_{23}-d_{12})/\lambda} \qquad (2)$$

where G_{ij} denotes the conductance between the corresponding layers.

The quantities θ and d_{ij} can be controlled with a large degree of precision. The twist angle θ can be tuned within $\sim 1°$ during fabrication [11], whereas d_{ij} can be varied by adding monolayers of dielectric materials, such as hBN or MoS_2. Since typical values $\lambda = \hbar/(2m_e W)^{1/2} \sim 2$ Å, estimated for the tunneling barrier height $W \sim 1$ eV and the effective electron mass $m_e \sim 10^{-30}$ kg, are comparable to the hBN or MoS_2 monolayer thickness, variation in d_{ij} results in a fairly gradual change of Z.

One appealing aspect of this system is the short interlayer transport length of a nanometer scale, which can allow high operation speeds and fast switching times. This is evident from an estimate for the RC time, $\tau_{RC} = \kappa/4\pi gd \sim 100$ ns, where $\kappa \sim 1$ is the dielectric constant, $d \sim 1$ nm is the interlayer separation, and $g \sim 10^{-7}\,\Omega^{-1}\mu\text{m}^{-2}$ is the interlayer conductance per unit area. The combination of geometric tunability and small transport lengths is not present in previously studied graphene-based bistable systems, such as graphene flash memories [2, 3] or graphene resistive memories [4–6]. Small thicknesses can also enable large packing densities.

The steep electronic dispersion in graphene makes the bistable state properties distinct from those in QW systems. In our case, the bistability is controlled by the resonances arising due to band alignment. The corresponding bias value, which scales as a power law of the energy Δ given in Eq.(1), can be as large as $\delta V \sim 100$-500 mV (see discussion below). In QW systems, instead, the bias range where bistability occurs is mainly controlled by the amount of charge n_{QW} that can be stored in a quantum well, $\delta V \approx en_{QW}/C$, where C is the interlayer capacitance [14]. Typical carrier densities in the "charged" and "uncharged" states of a bistable QW system, assessed by magnetic oscillation measurements[15], are on the order of $n_{QW} \sim 10^{11}/\text{cm}^2$ and $n_{QW} \sim 0$, respectively. These carrier densities yield typical values $\delta V \sim 50$ mV in double-barrier quantum wells with a width of tens of nanometers ($C \sim 0.1$-1 mF). Such values can be as much as an order of magnitude smaller than the above estimate predicts for the graphene case.

II. SEQUENTIAL TUNNELING MODEL

Vertical transport in our trilayer architecture can be described by a simple sequential model. The model validity relies on the interlayer tunnel coupling being weak such that the inter-layer charge transfer is slow compared to the intra-layer electron relaxation. Indeed, the values τ_{RC}, estimated above, are much longer than typical thermalization times in graphene, $\tau_{th} \sim 10$ ps [16]. The RC times, however, are sufficiently fast to be competitive with the speeds of existing switching devices [1].

The interlayer transport mechanism is mainly governed by the twist angle θ, which defines the K-point displacement $\mathbf{q}_A$ between graphene lattices in adjacent layers, and the interlayer bias. Under bias, the value $|\mathbf{q}_A|$ given in Eq.(1) determines the range of momenta and energies for which momentum-conserving tunneling is allowed. Large values of $|\mathbf{q}_A|$ hinders resonant tunneling given that phonon and defect scattering are necessary to supply the large momentum mismatch between layers. Momentum-nonconserving transport can also occur if the top/bottom layer is made of a different material so that there is a large mismatch between unit cells with respect to that of graphene. For small $|\mathbf{q}_A|$, on the other hand, momentum conserving tunneling is possible for moderately small values of bias.

In our two-step sequential tunneling model, we treat transport between layers 1 and 2 as momentum-conserving. The second step, between layers 2 and 3, is assumed to be momentum-nonconserving and described by Ohm's law. The latter assumption allows us to simplify our discussion and focus on the essential aspects of bistability. In addition, we also assume that the contact resistances are sufficiently small so that all the potential drop occurs predominantly between the graphene layers.

Turning to a systematic development of the model, the low energy Hamiltonian $\mathcal{H}$ describing coherent transport between a pair of twisted graphene monolayers has contributions $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2 + \mathcal{T}_{12}$. Here $\mathcal{H}_{1,2}$ are the free-particle terms describing massless Dirac particles in each graphene layer, and $\mathcal{T}_{12}$ describes the interlayer tunnel coupling [17–19]. The free particle terms are

$$\begin{aligned}
\mathcal{H}_1 &= \sum_{\mathbf{k}} \psi_{1,\mathbf{k}}^\dagger [\hbar v_F \boldsymbol{\sigma} \cdot (\mathbf{k} + \mathbf{q}_A/2) - \mu_1]\psi_{1,\mathbf{k}}, \\
\mathcal{H}_2 &= \sum_{\mathbf{k}} \psi_{2,\mathbf{k}}^\dagger [\hbar v_F \boldsymbol{\sigma} \cdot (\mathbf{k} - \mathbf{q}_A/2) - \mu_2]\psi_{2,\mathbf{k}},
\end{aligned} \qquad (3)$$

where $\mu_{1,2}$ are the Fermi energies measured relative to the Dirac point. For a small twist angle θ, the large-wavenumber processes that couple different valleys can be neglected. In this case, it is sufficient to account for a single Dirac cone in each layer, see Eq.(3). We adopt this approximation below.
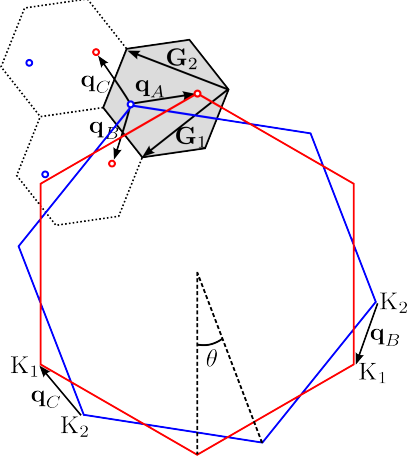
The tunneling coupling can be modeled as a local, pe-

FIG. 2. Twisted graphene layers form an hexagonal superlattice with reciprocal superlattice vectors $\mathbf{G}_1$ and $\mathbf{G}_2$ [17]. The momentum conserving tunneling coupling has the periodicity of the superlattice and can be decomposed into Fourier components $\mathbf{G} = n\mathbf{G}_1 + m\mathbf{G}_2$, with n, m being integers. For a small twist angle θ, tunneling is dominated by the smallest wavevectors $\mathbf{q}_A$, $\mathbf{q}_B = \mathbf{q}_A - \mathbf{G}_1$ and $\mathbf{q}_C = \mathbf{q}_A - \mathbf{G}_2$, see Eq.(3).

riodic function of position [17]:

$$\mathcal{T}_{12} = \sum_{\mathbf{k},\mathbf{G}} \psi_{1,\mathbf{k}}^\dagger \mathbf{T}_{\mathbf{G}} \psi_{2,\mathbf{k}+\mathbf{G}} + \text{H.c.}. \tag{4}$$

The periodicity of the interlayer coupling, quantified by the $\mathbf{G}$ wavevectors, is determined by the hexagonal superlattice unit cell that is formed by the twisted graphene layers, see Fig.2. For small θ, only the longest wavelength contributions are relevant for tunneling. Referred from the Dirac point of layer 1, such long wavelength components are given by $\mathbf{q}_A$, $\mathbf{q}_B = \mathbf{q}_A - \mathbf{G}_1$ and $\mathbf{q}_C = \mathbf{q}_A - \mathbf{G}_2$ (see Fig.2), where $\mathbf{G}_{1,2}$ are the reciprocal vectors of the superlattice Brillouin zone, which is smaller than the graphene Brillouin zone by a factor $\sim \sin^2(\theta)$. While the higher-$\mathbf{q}$ harmonics of the interlayer hopping potential spatial modulation also contribute to tunneling, it can be shown that their contributions vanish rapidly on the reciprocal lattice vector $\mathbf{G}_{1,2}$ scale [17, 18]. This leads to the tunneling Hamiltonian

$$\mathcal{T}_{12} = \sum_{j=A,B,C} \sum_{\mathbf{k}} \psi_{1,\mathbf{k}}^\dagger \mathbf{T}_j \psi_{2,\mathbf{k}+\mathbf{q}_j} + \text{H.c.} \tag{5}$$

comprised of only three Fourier components. In this expression for $\mathcal{T}_{12}$, the $\mathbf{k}$ vectors are relative to the Dirac point of each layer, i.e. $\mathbf{k} - \mathbf{q}_A/2 \to \mathbf{k}$ in layer 1 and $\mathbf{k} + \mathbf{q}_A/2 \to \mathbf{k}$ in layer 2.

Parenthetically, the lattice of the dielectric material separating the graphene layers can produce slowly varying spatial modulation of the tunneling transition amplitude $\mathbf{T}$ in Eq.(5), giving rise to the effects resembling those due to a twist angle θ. This would be the case when the dielectric and graphene are nearly lattice-matched as

e.g. in highly-oriented hBN-graphene structures, which have a small lattice mismatch of about 1.8% (a detailed discussion of these effects can be found in Ref.[20]). Such effects, if present, would alter the values $\mathbf{q}_{A(B,C)}$ but otherwise not change our discussion in an essential way.

Under an interlayer bias potential V_{12}, the tunneling current I_{12} is

$$\begin{aligned} I_{12} = \ & \frac{eN}{\hbar} \sum_{\mathbf{k}ss'j} |T_j^{ss'}(\mathbf{k})|^2 \int_{-\infty}^\infty \frac{d\omega}{2\pi} A_{1,s}(\mathbf{k},\omega) \\ & \times A_{2,s'}(\mathbf{k}+\mathbf{q}_j, \tilde\omega) \left[f_1(\omega) - f_2(\tilde\omega) \right], \end{aligned} \tag{6}$$

where s (s') refers to the electron ($+$) and hole ($-$) bands of layer 1 (2), and $N = 4$ is the spin and valley degeneracy. The functions $f_i(\omega) = 1/[e^{\beta(\omega-\mu_i)}+1]$ are the Fermi distribution functions for each layer, with $\beta = 1/k_\text{B}T$ being the inverse thermal energy and μ_i being the Fermi energies. The function $A_{i,s}$ is the spectral function of layer i and band s. The energy for the quantities in layer 2 is offset by $\tilde\omega = \omega + e\Phi_{12}$ due to the built-up interlayer electrostatic potential Φ_{12} [see Eq.(3)] between layers 1 and 2. Because of capacitance effects, the interlayer electrostatic and chemical potentials are related by

$$eV_{12} = \mu_1 - \mu_2 - e\Phi_{12}, \tag{7}$$

where μ_i and Φ_{12} are implicit functions of V_{12}. The quantity $T_j^{ss'}$ in Eq.(6) denotes

$$\begin{aligned} T_j^{ss'}(\mathbf{k}) &= \langle \mathbf{k}, s, 1 | \mathbf{T}_j | \mathbf{k} + \mathbf{q}_j, s', 2 \rangle, \\ |\mathbf{k}, s, i\rangle &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ se^{i\theta_\mathbf{k}} \end{pmatrix}, \end{aligned} \tag{8}$$

where $|\mathbf{k}, s, i\rangle$ are the two-component eigenvectors of $\mathcal{H}_{1,2}$ in Eq.(3) and $\theta_\mathbf{k}$ is the $\mathbf{k}$-vector polar angle.

The bistability can now be described by combining relations (3) and (7) as follows. In a steady state, there is zero net flow of carriers into the middle layer. Therefore, when the external bias $V = V_{12} + V_{23}$ between top and bottom layers is fixed, the equilibrium current I is obtained by solving for V_{12} from the non-linear equation

$$I(V) = I_{12}(V_{12}) = I_{23}(V - V_{12}). \tag{9}$$

This procedure to obtain the I-V response is shown graphically in the inset of Fig.1c. The straight line describes transport between layers 2 and 3 which is assumed to follow Ohm's law, $I_{23} = G_{23}V_{23}$, where G_{23} and V_{23} are the interlayer conductance and interlayer bias potential between layers 2 and 3, respectively.

III. ELECTROSTATIC FEEDBACK

In order to include the electrostatic feedback, Eq.(9) needs to be complemented with further electrostatic considerations that relate the variables V_{ij}, Φ_{ij}, and μ_i. It

is important to note that all variables can be determined once the carrier densities in each layer, n_1, n_2 and n_3, are known. Indeed, assuming that there is no external gate, the neutrality condition relates the charge densities in the different regions of the device as

$$n_1 + n_2 + n_3 = 0. \tag{10}$$

Furthermore, the application of an external bias potential V fixes the Fermi level difference between layer 1 and layer 3 as

$$eV = \mu_1 - \mu_3 + \frac{4\pi e^2}{\kappa}(n_1 d_{13} + n_2 d_{23}). \tag{11}$$

Here d_{ij} is the interlayer distance between layer i and layer j, κ is the dielectric constant of the barrier material, and $\mu_i = \text{sgn}(n_i)\hbar v_F \sqrt{\pi n_i}$. In Eq.(11), we implicitly assume that all layers are undoped at $V = 0$. Equations(9)-(11) then form a closed set of equations from which n_1, n_2 and n_3 can be obtained. The remaining variables, V_{ij} and Φ_{ij}, are functions of n_i. In particular, the electrostatic potentials are $\Phi_{12} = -4\pi e^2 d_{12} n_1/\kappa$ and $\Phi_{23} = 4\pi e^2(n_1 d_{13} + n_2 d_{23})/\kappa$, whereas the interlayer bias potentials are $V_{12} = \mu_1 - (\mu_2 + \Phi_{12})$, and $V_{23} = (\mu_2 + \Phi_{12}) - (\mu_3 + \Phi_{23})$.

For simplicity, here we fix the Fermi energies in Eq.(6) to a constant value $\mu_i = \mu$. This is equivalent to turning off capacitance effects. In this case, $V_{ij} = \Phi_{ij}$ (see Fig.3). This approximation is valid in the regime $4e^2 d_{ij}\Delta/\kappa(\hbar v_F)^2 \approx 15 \cdot d_{ij}[\text{nm}]\Delta[\text{eV}]/\kappa \gg 1$. In this regime, minimal changes in carrier concentration induce large interlayer electrostatic potentials. The more realistic scenario which includes quantum capacitance effects[21], such that $\mu_{1,2}$ vary with V_{12}, is here considered in Appendix A. However, this more realistic picture only introduces small corrections to the tunneling current without major consequences to our bistability discussion.

IV. MODEL PARAMETERS

In order to solve Eq.(9), we need to specify the matrix elements $\mathbf{T}_j$ in Eq.(8). A simple and explicit model for $\mathbf{T}_j$ and the wavevectors $\mathbf{q}_j$ is provided by Ref.[17]:

$$\mathbf{T}_j = t\begin{pmatrix} e^{i\varphi_j} & 1 \\ e^{-i\varphi_j} & e^{i\varphi_j} \end{pmatrix}, \quad \mathbf{q}_j = \frac{\Delta}{\hbar v_F}(\sin\varphi_j, -\cos\varphi_j), \tag{12}$$

with $\varphi_A = 0$, $\varphi_B = 2\pi/3$, $\varphi_C = 4\pi/3$. This representation is obtained for small twisting angles after performing a θ rotation of phase space in layer 2 (see details in Ref.[17]). It is also implicit in Eq.(12) that the top and bottom graphene lattices have a common lattice point [17]; a rigid horizontal translation between lattices adds an additional overall phase to the matrix $\mathbf{T}_j$ [18]. We stress, however, that relative phases in $\mathbf{T}_j$ do not alter

in any significant way the physics of tunneling in Eq.(6). Furthermore, while the interlayer hopping amplitude t is sensitive to several parameters, e.g. twist angle [19] and the choice of dielectric material [20], its order of magnitude is mainly governed by the wavefunction overlap between the graphene layers. Such dependence will be described below within the WKB approximation. Equations (3) and (12) are expected to be accurate for twist angles $\theta \lesssim 10°$, and energies of $1\,\text{eV}$ [19].

For an estimate below we use the value $\theta = 2°$. This defines an energy scale $\Delta = 0.37\,\text{eV}$. Furthermore, we take a Lorentzian spectral function in Eq.(6) for both layers, $A_{i,s}(\mathbf{k}, \omega) = 2\Gamma/\left[(\omega - s\hbar v_F|\mathbf{k}|)^2 + \Gamma^2\right]$ with the linewidth $\Gamma \sim 10\,\text{meV}$. A finite linewidth Γ is necessary to have a finite value of the peak current when $eV_{12} = \Delta$ (see Fig.3). The temperature and Fermi level of the system were taken to be $T = 0$ and $\mu_i = 0$, respectively. With reference to Eq.(6), we define the interlayer conductance

$$G_{12} = S g_{12}, \quad g_{12} = 2\pi N \frac{|t|^2}{(\hbar v_F)^2}\frac{e^2}{h}, \tag{13}$$

where S is the surface area of the device. The value of g_{12} is sensitive to the twist angle and the stacked dielectric material, if any, via the parameter t. Here we use $g_{12} = 10^{-7}\Omega^{-1}\mu\text{m}^{-2}$. Similar values of g_{12} were measured in resonant tunneling devices which contained four layers of BN in-between the graphene layers [12]. For Z, we consider a value of $Z = G_{12}/G_{23} = 0.2$.

V. BISTABLE I-V CHARACTERISTICS

The bistable I-V characteristics are shown in Fig.1c. For a sufficiently large bias, $eV \gtrsim \Delta$, the current branches into two stable states. The low-resistance branch in Fig.1c corresponds to two layers at resonance (i.e., $eV_{12} \approx \Delta$), whereas the high-resistance branch corresponds to a non-resonant state (i.e., $eV_{12} > \Delta$). We note that a third solution is also possible, indicated by a dashed line in the I-V response (see Fig.1c). This solution, however, is unstable given that a small perturbation in δV_{12} will push the system away from this state.

The bistable bias range can be estimated as $\delta V \approx (I_{12}^{(\text{pk})} - I_{12}^{(\text{vl})})/G_{23}$, where $I_{12}^{(\text{pk})}$ is the peak interlayer current and $I_{12}^{(\text{vl})}$ is the valley interlayer current (see inset of Fig.1c). To estimate $I_{12}^{(\text{pk})}$ and $I_{12}^{(\text{vl})}$, we first note that the tunneling matrix element $T_j^{ss'}(\mathbf{k})$ varies, upon integration in $\mathbf{k}$-space, in the range $0 \leq |T_j^{ss'}(\mathbf{k})| \leq 2t$ taking typical values $|T_j^{ss'}(\mathbf{k})| \approx t$. Thus, it is a good approximation to take band and wavevector-independent tunneling matrix elements $|T_j^{ss'}(\mathbf{k})| = \bar{T}$. Furthermore, in the typical case scenario the model parameters satisfy $\Gamma(\sim 10\,\text{meV}) \ll \Delta(\sim 0.1 - 1\,\text{eV})$. With this in mind, the integration of Eq.(6) allows an analytical expression
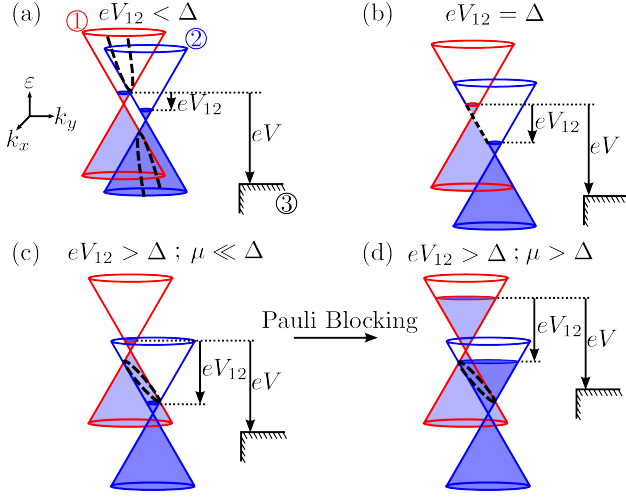
FIG. 3. Regions in **k**-space contributing to the resonant tunneling current for fixed V. These regions, indicated with black dashed lines at the intersection of the twisted Dirac cones, form conical surfaces in the **k**-plane: when $eV_{12} < \Delta$ the lines form hyperbolic curves, and when $eV_{12} > \Delta$ the lines form ellipsoidal curves. When $eV_{12} = \Delta$, a van-Hove singularity in the tunneling density of states is obtained. As shown in (d), the non-resonant (high-resistance) bistable state ($eV_{12} > \Delta$) can be Pauli-blocked by adjusting the doping. Doping thus affords a way to tune the current ratio between bistable branches in Fig.1a. In this work it is assumed that the Dirac cones are aligned at $V = 0$, and that capacitance effects are neglected. Layers are labeled 1-3 as in Fig.1a.

to result in terms of line integrals in conical surfaces (see Fig.3 and the discussion in Appendix A). Using $\mu_{1,2} = 0$ and $V_{12} = \Phi_{12}$, we find that the non-resonant interlayer current takes the simple form

$$\frac{I_{12}(x)}{I_{12}^{(\mathrm{vl})}} = \frac{x^2 - 1/2}{\sqrt{2(x^2 - 1)}}, \quad I_{12}^{(\mathrm{vl})} = \frac{3\sqrt{2}\bar{T}^2}{4}\frac{G_{12}\Delta}{e}. \quad (14)$$

Here $x = eV_{12}/\Delta \gtrsim 1$ and $I_{12}^{(\mathrm{vl})}$ is the valley current obtained at $x = \sqrt{3/2}$. When $eV_{12}/\Delta = 1$, however, the current is at resonance and reaches a maximum value which is sensitive to Γ. To leading order in Γ, we obtain (see Appendix B)

$$I_{12}^{(\mathrm{pk})}/I_{12}^{(\mathrm{vl})} = \pi\sqrt{\Delta/2\Gamma}, \quad (15)$$

where Γ, in general, depends on the amount and type of disorder and/or temperature. Equations (14) and (15) yield $e\delta V/\Delta \approx 3\sqrt{2}\bar{T}^2 Z[\pi\sqrt{\Delta/2\Gamma} - 1]/4$. Importantly, very small values of Z ($G_{23} \gg G_{12}$) make the bistable bias range negligibly small, whereas large values of Z ($G_{23} \ll G_{12}$) would push the onset of the bistability region to very large bias potentials. Optimally, values of $Z \sim 1$ and very small Γ would make the bistability effect more prominent.

Achieving a large current ratio between bistable states is desirable for applications; this facilitates the reading process in a bistable device. From Eqs.(14) and (15), it is obtained that the current ratio between bistable branches is controlled by the parameter $Z\sqrt{\Delta/\Gamma}$. For realistic values of disorder, this ratio can be in the 1-20 ballpark. It is interesting to note that these already high values can be boosted by means of Pauli blocking. As shown in Figs.3c and d, for sufficiently heavily doped samples, the non-resonant bistable state (but not the resonant one) is Pauli-blocked. The degree of the electrical current ratio enhancement depends on second order processes which assist tunneling, such as scattering with defects or disorder. These second order processes are not considered here.

The geometric control of Z, an appealing aspect of our system, can be understood from the Bardeen Transfer Hamiltonian Theory [22, 23]. In this theory, the interlayer coupling t is calculated from the overlap of the wavefunctions of layers i and j in the barrier region, $t = (\hbar^2/2m_e)\int d\mathbf{S} \cdot (\psi_i^* \nabla\psi_j - \psi_j \nabla\psi_i^*)$, with $d\mathbf{S}$ being a surface area element. Considering electrons tunneling across a square potential barrier with a height much larger than the electron kinetic energy, a tunneling matrix element of the form $t \propto \exp(-d_{ij}/\lambda)$ is obtained, where λ is the WKB decay length defined above. The expression of Z in Eq.(2) results from assuming barriers between layers 1 and 2 and between layers 2 and 3 are of the same material, in combination with Eq.(13).

Although electrostatic doping of the graphene layers is not essential for the physics that we describe, it is a convenient feature of bistability. In particular, for a fixed external bias potential, each bistable state exhibits different carrier concentrations. Thus, any in-plane measurement, such as conductance or magneto-transport, will be able to distinguish two distinct bistable states. Indeed, from the inset of Fig.1c we see that the interlayer bias potential for each bistable state differs by an amount $\delta V_{12} \sim \Delta/e$ (see also discussion in the Appendix A). Taking into account the capacitance of the layers, the induced carrier difference between both states is approximately $\delta n \sim \kappa\Delta/4\pi e^2 d_{12}$ (here the quantum capacitance is not included). Using $\theta = 2°$, $\kappa = 1$, and $d_{12} = 1\,\mathrm{nm}$, we obtain a carrier density difference $\delta n \sim 10^{12}\,\mathrm{cm}^{-2}$ between stable states. These large carrier density differentials can be used as a smoking gun of intrinsic bistability.

VI. OTHER GRAPHENE-BASED BISTABLE SYSTEMS

Although we considered here for simplicity a two-step sequential tunneling structure where only one pair of layers can be resonant, similar ideas apply to more complex structures. Interesting examples include a two-step resonant-resonant structure, opening the possibility for tristability or multi-step "cascade" devices.

Finally, we also expect bistable I-V characteristics in twisted graphene trilayers in the absence of any di-

electric material. Indeed, incommensurability between graphene lattices already suppresses interlayer hybridization, regardless of the layers being spatially separated by a fraction of a nanometer, thus enabling the sequential tunneling regime [18]. Furthermore, the massless Dirac spectrum, and thus Eq.(3) and the subsequent transport model, remain valid but with a modified Fermi velocity [17]. We stress, however, that stacked dielectric materials have two important advantages: (i) they enable tuning the interlayer coupling and (ii) they facilitate the interlayer potential build-up in order to achieve a resonant behavior.

VII. SUMMARY

In summary, graphene-based van der Waals heterostructures afford a new platform to realize devices with tunable I-V characteristics, in particular those with intrinsically bistable and hysteretic behavior. System parameters required to realize the bistable behavior are readily accessible in on-going experiments. The atomic scale interlayer distances can result in a fast response and large packing-densities, making these heterostructures appealing for a variety of applications.

ACKNOWLEDGEMENTS

[1] H.-S. P. Wong and S. Salahuddin, Nat Nanotechnol. **10**, 191 (2015).
[2] A. J. Hong, E. B. Song, H. S. Yu, M. J. Allen, J. Kim, J. D. Fowler, J. K. Wassei, Y. Park, Y. Wang, J. Zou, R. B. Kaner, B. H. Weiller, and K. L. Wang, ACS Nano **5**, 7812 (2011).
[3] S. Bertolazzi, D. Krasnozhon, and A. Kis, ACS Nano **7**, 3246 (2013).
[4] X. Wang, W. Xie, J. Du, C. Wang, N. Zhao, and J.-B. Xu, Adv. Mater. **24**, 2614 (2012).
[5] Y. J. Shin, J. H. Kwon, G. Kalon, K.-T. Lam, C. S. Bhatia, G. Liang, and H. Yang, Appl. Phys. Lett. **97**, 262105 (2010).
[6] Y. Zheng, G.-X. Ni, C.-T. Toh, M.-G. Zeng, S.-T. Chen, K. Yao, and B. Özyilmaz, Appl. Phys. Lett. **94**, 163505 (2009).
[7] D. B. Strukov, G. S. Snider, D. R. Stewart and R. S. Williams, Nature **453**, 80 (2008).
[8] J. J. Yang, D. B. Strukov and D. R. Stewart, Nat Nanotechnol **8**, 13 (2013).
[9] A. D. Kent, and D. Worledge, Nat Nanotechnol **10**, 187 (2015).
[10] A. K. Geim and I. V. Grigorieva, Nature **499**, 419 (2013).
[11] A. Mishchenko, J. S. Tu, Y. Cao, R. V. Gorbachev, J. R. Wallbank, M. T., Greenaway, V. E. Morozov, S. V. Morozov, M. J. Zhu, S. L. Wong, F. Withers, Y-J. Woods, C. R. Kim, K. Watanabe, T. Taniguchi, E. E. Vdovin, O. Makarovsky, T. M. Fromhold, V. I. Fal'ko, A. K. Geim, L. Eaves, and K. S. Novoselov, Nature Nano **9**, 808 (2014).
[12] L. Britnell, R. V. Gorbachev, A. K. Geim, L. A. Ponomarenko, A. Mishchenko, M. T. Greenaway, T. M. Fromhold, K. S. Novoselov, and L. Eaves, Nature Commun. **4**, 1794 (2013).
[13] V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Phys. Rev. Lett. **58**, 1256 (1987).
[14] F. W. Sheard and G. A. Toombs, Appl. Phys. Lett. **52**, 1228 (1988).
[15] A. Zaslavsky, V. J. Goldman, D. C. Tsui, and J. E. Cunningham, Appl. Phys. Lett. **53**, 1408 (1988).
[16] P. A. George, J. Strait, J. Dawlaty, S. Shivaraman, M. Chandrashekhar, F. Rana and M. G. Spencer, Nano Lett. **8**, 4248 (2008).
[17] J. M. B. Lopes dos Santos, N. M. R. Peres, and A. H. Castro Neto, Phys. Rev. Lett. **99**, 256802 (2007).
[18] R. Bistritzer and A. H. MacDonald, Phys. Rev. B **81**, 245412 (2010).
[19] R. Bistritzer and A. H. MacDonald, PNAS **108**, 12233 (2011).
[20] L. Brey, Phys. Rev. Appl. **2**, 014003 (2014).
[21] S. Luryi, Appl. Phys. Lett. **52**, 501 (1988).
[22] J. Bardeen, Phys. Rev. Lett. **6**, 57 (1961).
[23] E. L. Wolf, *Principles of Electron Tunneling Spectroscopy*, Second Edition (Oxford University Press, Oxford, 2012)

Appendix A: Capacitance effects

In the main text, we fixed the Fermi energy μ_i of the different graphene layers to some constant value. A more refined model of the I-V response should, however, include quantum capacitance effects so that Fermi energy is allowed to vary with V. Although the features of bistability are not significantly modified by such corrections, as shown below, carrier density differentials between the bistable states are a smoking gun of intrinsic bistability. These electrostatic considerations are discussed next.

Here we numerically solve Eqs.(6)-(11), assuming a thin device separated by dielectric barriers of thickness $d_{12} = d_{23} = 1.4\,\text{nm}$ (e.g. four layers of hBN) and dielectric constant $\kappa = 5$. The procedure to solve the I-V response self-consistently is shown in Fig.A1a, where n_1 and n_2 are taken as independent variables [n_3 is obtained from Eq.(10)], and $\delta I = I_{12} - I_{23}$ in Eq.(6) is numerically calculated (color map). For fixed V, indicated with dotted isolines in Fig.A1a, the self-consistent solutions to
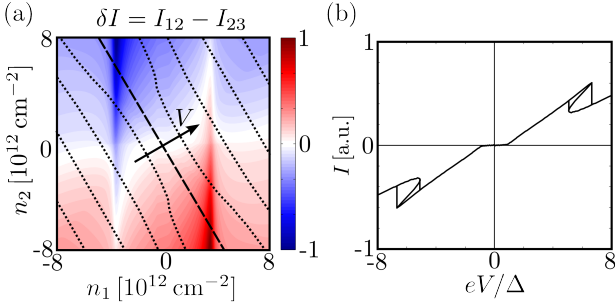
FIG. A1. Self-consistent bistable solutions including quantum capacitance effects [Eqs.(10)–(11)]. For fixed V, we find n_1 and n_2 such that $\delta I = I_{12} - I_{23} = 0$. The bias isolines from Eq.(11) are marked with dashed ($V = 0$) and dotted (finite V) lines, with an arrow pointing towards increasing V. The self-consistent I-V curve, obtained from the intersection of $\delta I = 0$ and the V-isolines in (a), is plotted in panel (b).

the equilibrium equations are given by the pair (n_1,n_2) such that $\delta I = 0$.

The resulting I-V response is shown in Fig.A1b. Importantly, the I-V characteristics are qualitatively similar to those obtained in the main text by neglecting quantum capacitance effects. Furthermore, by inspection of the n_1 and n_2 axes in Fig.A1a, we see that the difference in carrier concentration δn between each bistable state is on the order of $\delta n \sim 10^{11}$-10^{12} cm^{-2}. These carrier concentration differences can easily be detected by lateral transport measurements and may act as clear fingerprints of intrinsic bistability.

Appendix B: Analytic expressions for the peak and valley resonant tunneling current

We derive here Eqs. (14) and (15) of the main text, obtained under the assumption that the tunneling matrix elements $T_j^{ss'}$ in Eq.(8) are independent of the wavevector and band index, i.e., $|T_j^{ss'}(\mathbf{k})| = \bar{T}$. Under this assumption, I_{12} depends only on the modulus of $\mathbf{q}_j$ but not on its direction, and $\sum_j |T_j^{ss'}(\mathbf{k})|^2 = 3\bar{T}^2$. Given that $\Gamma \ll \Delta$, when $e\Phi_{12} > \Delta$ (non-resonant state) we can set $\Gamma \to 0$ and thus take $A_{i,s}(\mathbf{k},\omega) = 2\pi\delta(\omega - s\hbar v_F|\mathbf{k}|)$. The two δ-functions appearing in Eq.(6) can then be integrated in $\mathbf{k}$-ω space, resulting in a one-dimensional integral along

the contour of an ellipse:

$$\sum_{ss'} \int \frac{d\mathbf{k}}{(2\pi)^2} \int_{\omega_1}^{\omega_2} \frac{d\omega}{2\pi} \delta(\omega - s\hbar v_F|\mathbf{k}|)\delta(\tilde{\omega} - s'\hbar v_F|\mathbf{k}+\mathbf{q}|)$$
$$= \frac{\delta_{s,-}\delta_{s',+}}{16\pi^3(\hbar v_F)^2} \int_{\phi_1}^{\phi_2} d\phi \frac{(e\Phi_{12})^2 - \Delta^2\sin^2\phi}{\sqrt{(e\Phi_{12})^2 - \Delta^2}}.$$
(A1)

Here we denote $\tilde{\omega}$ by $\tilde{\omega} = \omega + e\Phi_{12}$. In addition, the limits of integration on ω are given by $\omega_1 = e\Phi_{12} + \mu_2$ and $\omega_2 = \mu_1$, whereas the limits of integration on ϕ are

$$\phi_i = \begin{cases} \pi/2, & x_i > 1 \\ \sin^{-1}(x_i), & -1 < x_i < 1 \\ -\pi/2, & x_i < -1 \end{cases}, \quad x_{1,2} = \frac{2\mu_{1,2} \pm e\Phi_{12}}{\Delta}.$$
(A2)

In obtaining Eq.(A1), we parametrized $\mathbf{k}$-space using coordinates $k_x = k_r\sin\phi/2$ and $k_y = \sqrt{k_r^2 - q^2}\cos\phi/2$, with $\mathbf{q}$ conveniently aligned in the x-direction. The integration over k_r absorbs the first δ function, setting $k_r = e\Phi_{12}/\hbar v_F$. Integration over ω absorbs the second δ function, fixing the limits of integration $\phi_{1,2}$ in Eq.(A2). Importantly, because $\Phi_{12} > \Delta$, the two δ-functions in Eq.(A1) can only be non-zero simultaneously when $s = -$ and $s' = +$ (i.e. holes of layer 1 tunnel into electronic states of layer 2, see Fig.3). Using $\mu_{1,2} = 0$ and $V_{12} = \Phi_{12}$, Eqs.(A1) and (A2) result in Eq.(14).

When $e\Phi_{12} = \Delta$, it is necessary to restore the finite linewidth to the Lorentzian spectral function $A_{i,s}(\mathbf{k},\omega) = 2\Gamma/\left[(\omega - s\hbar v_F|\mathbf{k}|)^2 + \Gamma^2\right]$. In this case, the integral for the tunneling current yields

$$\sum_{ss'} \int \frac{d\mathbf{k}}{(2\pi)^2} \int_{\Phi_{12}+\mu_2}^{\mu_1} \frac{d\omega}{2\pi} A_{1,s}(\mathbf{k},\omega)A_{2,s'}(\mathbf{k}+\mathbf{q},\omega) =$$
$$= \frac{2}{(\hbar v_F)^2\sqrt{\Gamma\Delta}} \left[\int_{\mu_2}^{\Delta+\mu_1} d\omega\, |\omega(\omega - \Delta)|^{1/2} + \mathcal{O}(\Gamma/\Delta)\right].$$
(A3)

In obtaining Eq.(A3), we transformed the integral of the spectral functions into a dimensionless integral of the form $I_{\text{res}}(\epsilon) = \int d^2\mathbf{x}\{[f(\mathbf{x})^2 + \epsilon][g(\mathbf{x})^2 + \epsilon]\}^{-1}$. The functions f and g satisfy $f(0) = g(0) = 0$ and have a null Jacobian $\det[\partial_\mathbf{x}f, \partial_\mathbf{x}g](0) = 0$ (here $\epsilon = \Gamma/\Delta$). It can be shown that $I_{\text{res}} \propto \epsilon^{-1/2}$, when $\epsilon \ll 1$. An expansion to leading order in powers of ϵ gives Eq.(A3). Setting $\mu_{1,2} = 0$ in Eq.(A3), the peak current $I_{12}^{(\text{pk})}$ in Eq.(15) of the main text is obtained.